



**Manchester
Metropolitan
University**

Bones, Oliver ORCID logoORCID: <https://orcid.org/0000-0002-1608-3459>
and Plack, Christopher J (2015) Subcortical representation of musical dyads:
Individual differences and neural generators. Hearing Research, 323. pp. 9-
21. ISSN 0378-5955

Downloaded from: <https://e-space.mmu.ac.uk/623249/>

Version: Published Version

Publisher: Elsevier

DOI: <https://doi.org/10.1016/j.heares.2015.01.009>

Usage rights: Creative Commons: Attribution 4.0

Please cite the published version

<https://e-space.mmu.ac.uk>



Research paper

Subcortical representation of musical dyads: Individual differences and neural generators



Oliver Bones*, Christopher J. Plack

School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK

ARTICLE INFO

Article history:

Received 28 July 2014

Received in revised form

7 January 2015

Accepted 19 January 2015

Available online 28 January 2015

ABSTRACT

When two notes are played simultaneously they form a musical dyad. The sensation of pleasantness, or “consonance”, of a dyad is likely driven by the harmonic relation of the frequency components of the combined spectrum of the two notes. Previous work has demonstrated a relation between individual preference for consonant over dissonant dyads, and the strength of neural temporal coding of the harmonicity of consonant relative to dissonant dyads as measured using the electrophysiological “frequency-following response” (FFR). However, this work also demonstrated that both these variables correlate strongly with musical experience. The current study was designed to determine whether the relation between consonance preference and neural temporal coding is maintained when controlling for musical experience. The results demonstrate that strength of neural coding of harmonicity is predictive of individual preference for consonance even for non-musicians. An additional purpose of the current study was to assess the cochlear generation site of the FFR to low-frequency dyads. By comparing the reduction in FFR strength when high-pass masking noise was added to the output of a model of the auditory periphery, the results provide evidence for the FFR to low-frequency dyads resulting in part from basal cochlear generators.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

“Consonance” refers to the pleasant and stable sensation produced when two or more musical notes are presented simultaneously in simple frequency ratios. McDermott et al. (2010) demonstrated that individual preference for consonant over dissonant two-note musical chords (“dyads”) correlated with preference for harmonicity (the closeness of fit of a series of frequency components to a harmonic series) over inharmonicity, suggesting that the perception of consonance is driven by the perceived harmonicity of the dyad.

Pitch may be represented in the brainstem due to the tendency of neurons to synchronize their firing to a particular phase of the basilar membrane (BM) vibration (“phase locking”; Brugge et al., 1969; Rose et al., 1971). The sustained phase locked response of populations of neurons at this stage of the auditory pathway can be measured as the

frequency-following response (FFR; for a review of anatomical generators see Krishnan, 2007), a scalp recorded auditory evoked potential which takes its name from the characteristic peaks in the waveform at periods corresponding to frequency components of the stimulus. Recent work suggests a relation between the integrity of the temporal coding represented by the FFR and pitch discrimination (Carcagno and Plack, 2011; Clinard et al., 2010; Krishnan et al., 2012; Marmel et al., 2013), musical experience (Bidelman et al., 2011a, 2011b; Wong et al., 2007), tone language experience (Krishnan et al., 2005), and the perception of musical consonance (Bidelman and Krishnan, 2009; Bones et al., 2014).

Bones et al. (2014) demonstrated that individual differences in consonance perception for dyads could be predicted by the representation of harmonicity in the spectrum of the FFR. By subtracting the FFR to each stimulus presented in its original onset polarity from the FFR to the stimulus presented with the onset polarity inverted, phase locking to the cochlear envelope was suppressed whilst phase locking to temporal fine structure (TFS) was enhanced (Goblick and Pfeiffer, 1969). When the FFR was processed this way young normal hearing participants with a stronger representation of the harmonicity of consonant relative to dissonant dyads in the spectrum of the FFR (“neural consonance index”; NCI) had a

* Corresponding author. Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, G36, Leung Kau Kui Building, Shatin, New Territories, Hong Kong. Tel.: +852 39433779.

E-mail address: olly@cuhk.edu.hk (O. Bones).

stronger preference for consonant over dissonant dyads. This suggests that temporal coding of the frequency components of the combined spectrum of two notes may be a mechanism for encoding the harmonicity, and consonance, of musical dyads. However, musical experience is also strongly associated with consonance preference (Bones et al., 2014; McDermott et al., 2010) and harmonicity preference (McDermott et al., 2010). Bones et al. found that the correlation between NCI and consonance preference did not remain significant when the effects of musical experience were controlled, suggesting that the relation between integrity of the representation of harmonicity in the brainstem and consonance preference could be driven by a codependence on musical experience. One purpose of the current study was to address the outstanding question of whether harmonicity of temporal coding in the brainstem as measured by the FFR predicts variation in consonance preference in individuals without musical experience. If so, this would support the hypothesis that consonance is associated with neural temporal coding, rather than with some other aspect of processing that might co-vary with musical experience.

Since the upper limit of phase locking in the inferior colliculus is approximately 2000 Hz (Krishnan, 2007), the stimuli used to measure the FFR are typically below this frequency. Another issue in this area that is currently unresolved is that of which region of the cochlea is represented by the FFR to low-frequency dyads. The response to a low-frequency pure tone at low to moderate levels (approximately < 50 dB above hearing threshold; Krishnan, 2007) is likely to be generated in the region of the cochlea with characteristic frequencies (CFs) close to the frequency of the tone, and is measurable in listeners with high-frequency hearing loss (Moushegian et al., 1978; Yamada et al., 1977). However, the suprathreshold FFR to a low-frequency tone is reduced in amplitude by high-pass masking noise above the frequency of the tone (Bledsoe and Moushegian, 1980; Davis and Hirsh, 1976; Gardi and Merzenich, 1979; Huis in't Veld et al., 1977). Gardi and Merzenich (1979) found that the response to a 500 Hz tone presented at a high level (100 dB SPL) was reduced in amplitude by approximately 50% by 60 dB SPL high-pass (2000 Hz) masking noise, suggesting that the FFR may have been at least partly generated basally to the region of the BM tuned to the tone. This reduction presumably represents the desynchronization (Marsh et al., 1972) of high-frequency neurons which in the absence of the masker had been phase locked to the stimulus frequency. More recently, it has been demonstrated that the FFR to a low-frequency tone is reduced in amplitude when preceded by a high-frequency tone, implying that neural adaptation of neurons in a high-frequency channel can attenuate the response to a low-frequency tone (H. Gockel, personal communication, May 22, 2014). Again, this suggests that the FFR is generated basally to the region tuned to the tone. This can be partly understood as a consequence of the frequency response of the basal region of the BM, which is reflected in the tuning of auditory nerve fibres: high-frequency fibres have sharp peaks at their best frequencies and steep high-frequency sides but broad low-frequency tails (Geisler et al., 1974; Kiang and Moxon, 1974; Rose et al., 1971). High level tones such as the 100 dB SPL tone used by Gardi and Merzenich (1979) cause the BM response to broaden and the neural response of fibres tuned to the tone to become saturated. As a consequence fibres innervating the basal side of the BM response come to dominate the neural response (de Boer, 1977). Dau (2003) suggested that displacement of the basal region of the BM may be necessary to produce a well-defined FFR. Dau used an auditory nerve (AN) model (Heinz et al., 2001) to simulate the FFR to a 300 Hz pure tone. When only 100–1500 Hz channels were included in the model the frequency of the tone was poorly represented by the simulated FFR, and the waveform was low in amplitude up to a stimulus level of 100 dB SPL. However,

when only 1500–10,000 Hz channels were included the simulated FFR had a periodic response to a stimulus of 80 dB SPL, increasing in amplitude as the tone was increased in level to 100 dB SPL. When high-frequency channels were used the synchrony of the modelled BM response in the base of the cochlea (e.g. see Dau et al., 2000; Shore and Nuttall, 1985) meant that the spikes in the model's nerve fibres were well aligned. The frequency of the stimulus was therefore well represented in the FFR simulated from the pooled response. When low-frequency channels were used the nerve fibres responded asynchronously, due to the asynchronous response of the BM at the apex of cochlea, resulting in a poorly defined FFR.

It is likely that the FFR to high-intensity low-frequency pure tones is at least partly generated basal to the region of the BM tuned to the tone. However, whilst the cochlear origin of the FFR for pure tones has been well researched, that for complex tones has not. A second purpose of the current study was to test the hypothesis that the FFR to 80 dB SPL low-frequency musical dyads originates from a portion of the cochlea basal to that tuned to the dyads. To address this, the effects of high-pass masking noise on the FFR were compared to the output of an auditory model.

2. Methods

2.1. Participants

All participants had normal hearing (absolute thresholds of 20 dB HL or better at octave frequencies between 500 and 8000 Hz). Thirteen (10 females) participated in the psychophysical section of the study (20–28 yrs, mean 23 yrs). Ten (7 females) participated in both the psychophysical and the electrophysiological part of the study (20–28 yrs, mean 24 yrs). All participants self-reported as having less than one year of experience learning to play a musical instrument, with that period having ceased at least five years prior to the study. All participants provided written informed consent in compliance with a research protocol approved by the University of Manchester Research Ethics Committee.

2.2. Psychophysical testing

2.2.1. Stimuli

Stimuli were diotic musical dyads, consisting of a low (“root”) note, and a high (“interval”) note. Root notes were the eight notes above and including the musical note D3, all taken from the equal temperament scale: D (146.83 Hz), D# (155.56 Hz), E (164.81 Hz), F (174.61 Hz), F# (185.00 Hz), G (196.00 Hz), G# (207.65 Hz), and A (220.00 Hz). Each root note was used to create two types of dyad; a consonant Perfect 5th and a dissonant Tritone. Dyads are named after the ratio of the fundamental frequencies (F0s) of the two notes (the size of the interval), approximately 3:2 in the case of the Perfect 5th and $\sqrt{2}$:1 in the case of the Tritone. The simple frequency ratio of the Perfect 5th results in the combined spectrum of the two notes having a strong harmonicity, whilst the complex frequency ratio of the Tritone results in an inharmonic spectrum (Fig. 1A–B). Dyads were low-pass filtered at 2500 Hz, chosen so as to be within the frequency response region of the headphones, while allowing for a spectral region to add the noise. The harmonics of each note were equal amplitude so that the overall level of each dyad was 80 dB SPL. Each dyad was 2 s in duration including 10 ms raised-cosine onset and offset ramps.

For completeness, masking noise was also used in the behavioural section of the study. Each dyad was presented in three conditions: with no masking noise (No Noise); with Gaussian noise with a 31 dB spectrum level (Level 1), band-pass filtered between 2600 and 7000 Hz; and with the same masking noise with a 41 dB spectrum level (Level 2). When presented, masking noise was

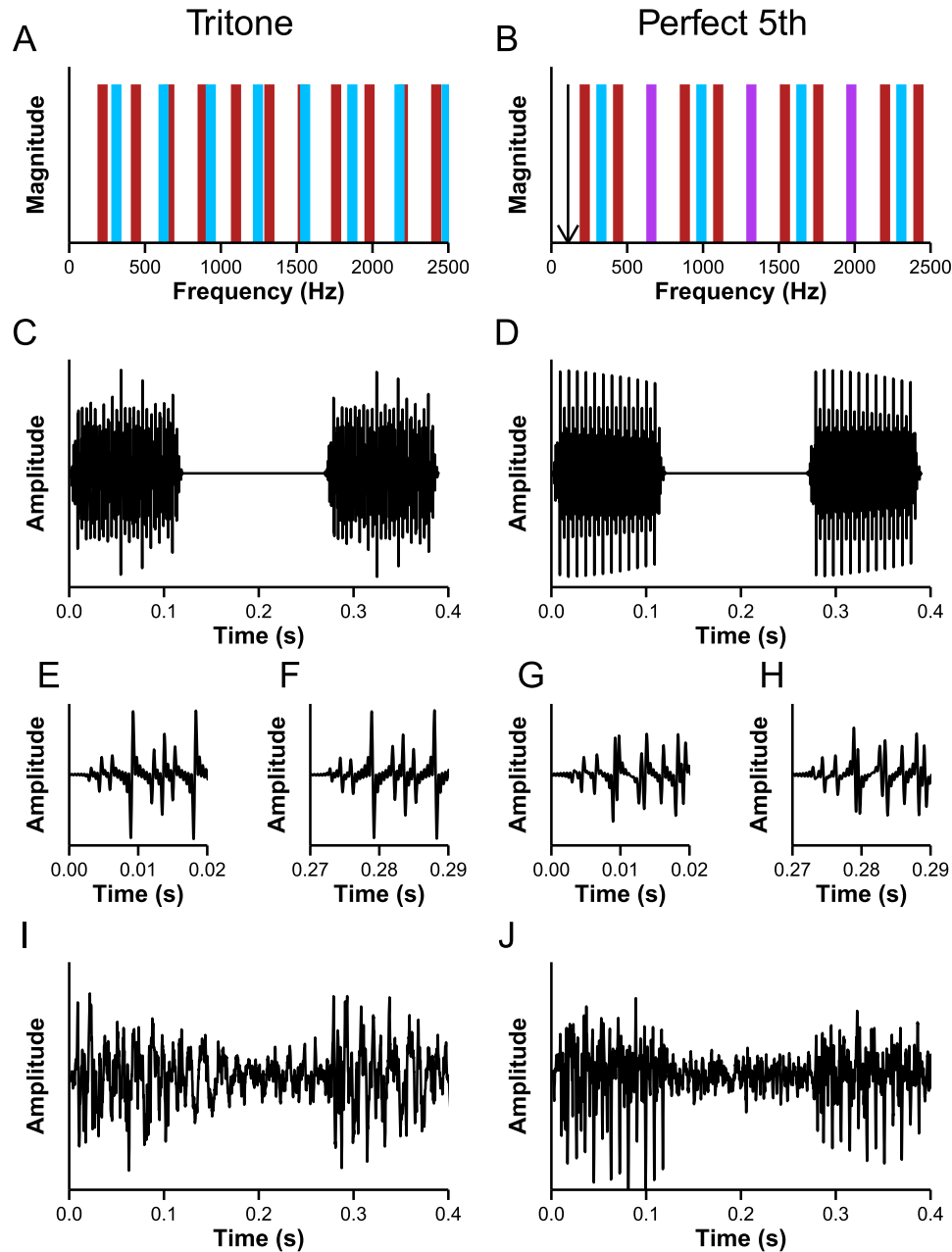


Fig. 1. The combined spectra of the root note (red) and the interval note (blue) of the Tritone (A) and the Perfect 5th (B). Where harmonics perfectly coincide they are coloured violet. The combined spectra of the Perfect 5th forms a harmonic series with an implied F0 indicated by the arrow. Dyads were presented in pairs during FFR recording, separated by 150 ms silence (C–D). The second dyad in each pair had a starting phase inverted 180° relative to the first dyad. To demonstrate this the first 20 ms of the Tritone (E–F) and the Perfect 5th (G–H) stimuli presented in each polarity are displayed. The average FFR waveform to the Tritone (I) is less periodic than the FFR to the Perfect 5th (J). The periodic peaks of the Perfect 5th stimulus waveform (D) can be identified in the FFR (J). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presented simultaneously and had the same duration and ramps as the dyads. The spectrum levels of the masking noise were chosen to be 15 and 5 dB respectively below the 46 dB spectrum level of the dyads, which had bandwidths of 2500 Hz when low-pass filtered.

Each stimulus was preceded by Gaussian noise with the same duration and ramps as the stimulus. This noise was low-pass filtered at 2500 Hz so that it had the same frequency range as the dyads. A 500 ms silence separated the noise and the stimulus. The purpose of the noise was to break up the sequence of the dyads to prevent any melodic structure from influencing responses (Bones et al., 2014; McDermott et al., 2010).

All stimuli were generated digitally in Matlab at a sampling rate of 24,414 Hz with a 32 bit resolution. Stimuli were delivered via a 24-bit E-MU 0202 USB audio device and Sennheiser HD 650 supra-aural headphones.

2.2.2. Procedure

Behavioural testing was performed first. Behavioural pleasantness ratings were measured by following the methodology of McDermott et al. (2010). Participants rated each dyad for pleasantness from −3 (very unpleasant) to +3 (very pleasant). Stimuli were presented in a random sequence. Participants were seated in a

sound attenuating booth and made responses via a PC keyboard and a monitor placed outside the booth, visible through a window. Each run consisted of 48 presentations (eight root notes \times two interval notes \times three masking noise conditions). Each participant performed two runs consecutively on the same day, preceded by one practice trial of one of each interval in each masking noise condition.

2.2.3. Analysis

Ratings for each interval (Perfect 5th and Tritone) in each masking noise condition were averaged across runs and root notes, so that each participant's rating for each interval in each masking noise condition was the mean of 32 responses. Individual consonance preference for each masking noise condition was calculated using a version of the routine described by McDermott et al. (2010): averaged pleasantness ratings were first z-scored for each individual in order to remove the influence of individual differences in the use of the scale. Consonance preference in each condition was then calculated by subtracting each individual's z-scored rating of the Tritone (a dissonant interval) from their z-scored rating of the Perfect 5th (a consonant interval).

2.3. Frequency-following responses

2.3.1. Stimuli and recording protocol

Stimuli were a subset of those used in psychophysical testing, without the intervening noise: diotic Perfect 5th and Tritone dyads with root notes set at A (220 Hz), with the same three masking noise conditions used in the psychophysical testing. There were therefore six stimuli in total (two dyads \times three masking noise conditions). Stimuli were 120 ms in duration, including 10 ms raised-cosine onset and offset ramps.

Electrodes were positioned at the high forehead hairline (active), the seventh cervical vertebra (reference), and at Fpz (ground; Bidelman and Krishnan, 2009; Bones et al., 2014; Gockel et al., 2011; Gockel et al., 2012; Krishnan and Plack, 2011; Krishnan et al., 2005). This electrode configuration allowed for direct comparison with Bones et al. (2014), and also ensured that contamination by the cochlear microphonic cannot occur (e.g. see Davis and Britt, 1984). Impedances were maintained below 3 k Ω . Participants were seated in a comfortable chair within a sound attenuating booth and told to remain as still as possible, with their arms, legs and neck straight, and told that they could sleep. Stimuli were delivered via a TDT RP2.1 Enhanced Real Time Processor and HB7 Headphone Driver and Etymotic ER3A insert earphones. Prior to each recording, a test recording was made with the headphone tubing clamped so that the participant could not hear the stimulus. Test recordings were then checked to ensure electromagnetic stimulus artefacts transmitted by the transducers were below the level of the background EEG.

FFRs to each dyad were collected separately. Presentations consisted of two dyads separated by 150 ms (Fig. 1C–D). Presentations were made at a rate of 1.82/s. The starting phase of the second of each pair of dyads was inverted 180° relative to the first presentation (Fig. 1E–H). The acquisition window lasted for 447 ms from the onset of the first presentation. Responses to each dyad were collected from each participant using TDT BioSig software. Responses were band-pass filtered online between 50 Hz and 3000 Hz. Data were compiled online as 200 sub-averages of 10 different responses to each stimulus polarity (2000 responses in total were averaged for each polarity, 4000 in total). Any sub-average in which the peak amplitude exceeded ± 15 μ V was assumed to be an artefact and rejected off-line. The artefact rejection threshold was chosen to achieve the lowest possible RMS (0.19 μ V) in the portion of the response between stimulus

polarities. An average of 13 sub-averages was rejected for each participant. The averaged waveforms of the FFR to the Tritone and the Perfect 5th with no masking noise are displayed in Fig. 1 I–J.

2.3.2. Analysis

A neural harmonic salience measure was used to determine how well the harmonicity of each dyad was represented by phase locking in the brainstem (Bones et al., 2014). The best fitting harmonic series to the power spectrum of each dyad was determined by finding the harmonic series for which the ratio of power inside 15 Hz wide bins placed at integer multiples of the F0 to power outside these bins was highest. Only F0s above 30 Hz were considered, since this was assumed to be the lower limit of pitch (Pressnitzer et al., 2001). The F0s of the best fitting harmonic series to the Perfect 5th and the Tritone were 55.00 and 44.25 Hz respectively. The salience of these harmonic series in the power spectrum of the FFR to each stimulus was then calculated in the same way, i.e. the ratio of power inside the bins placed at integer multiples of the F0 to power outside the bins.

Because the FFR was recorded to each stimulus in alternating polarity, the spectra upon which analyses were performed were the mean spectra of the responses to each polarity i.e. the power spectra, which do not contain phase information, of the responses to each polarity were calculated separately and then averaged together. This is referred to hereafter as FFR_{RAW}. In order to also investigate differences between the results of the current study and results reported previously (Bones et al., 2014), the same method for enhancing or suppressing phase locking to the envelope or TFS of the cochlear response was used: the mean FFR waveform to the original polarity was added to the mean FFR waveform to the inverted polarity to enhance the envelope response and suppress the TFS response (FFR_{ADD}); and the mean FFRs were subtracted to suppress the envelope response and enhance the TFS response (FFR_{SUB}; Gobblick and Pfeiffer, 1969).

In an abbreviated version of the NCI measure described elsewhere (Bones et al., 2014), the strength of the representation of harmonicity in the FFR to the Perfect 5th relative to the Tritone was calculated as the neural harmonic salience of the former minus the neural harmonic salience of the latter. The NCI measure is a physiological analogue to the consonance preference measure (see section 2.2.3).

2.4. Auditory model

The FFR data were compared to the output of a simulation of the auditory periphery using the Development System for Auditory Modelling library (DSAM; <http://dsam.org.uk/>). The same Perfect 5th stimulus files that were used to measure the FFR were used as inputs to the model.

The first stage of the simulation was an outer/middle ear filter model. This consisted of two parallel cascades of first-order band-pass filters. The first cascade consisted of two filters with 3 dB down points at 4000 and 25,000 Hz, the second consisted of three filters with 3 dB down points at 700 and 30,000 Hz. The output was then converted to stapes velocity using the scalar 1.4×10^{-10} (Sumner et al., 2003).

The BM was modelled using a dual resonance nonlinear (DRNL) filter (Meddis et al., 2001). In order to determine whether the FFR was best accounted for by the apical region of the cochlea tuned to the dyad or by a basal region, two models were used: a low-frequency model (LF model) only included the output of 10 filter channels between 100 and 2500 Hz, spaced using the Greenwood CF spacing function; and a high-frequency model (HF model) only

included the output of 10 cochlear filter channels between 2500 and 8000 Hz, spaced the same way.

The multi-channel output of the BM model was then used as the input to an inner hair cell receptor potential (IHC RP) simulation (Shamma et al., 1986) using parameters provided by Sumner et al. (2002). The output was then averaged across frequency channels and used to simulate the FFR.

2.5. Statistics

The effect of musical interval (Tritone and Perfect 5th) and level of noise masking (No Noise, Level 1, Level 2) on pleasantness ratings and neural harmonic salience were examined using repeated measures ANOVAs. Generalized Eta-Squared effect sizes are reported (η^2). Where the assumption of sphericity was violated, the Greenhouse–Geisser method was used and correction factors are reported (ϵ). Post-hoc pairwise comparisons were corrected using the Bonferroni method. Correlations were tested using Spearman's Rho test.

3. Results

3.1. The effect of high-pass masking noise on pleasantness ratings of low-frequency dyads

Pleasantness ratings for each interval in each masking noise condition are plotted in Fig. 2A. For both intervals variance in rating was greater with no masking noise (Tritone, $SD = 1.31$; Perfect 5th, $SD = 1.23$) than with Level 1 (Tritone, $SD = 0.61$; Perfect 5th, $SD = 0.80$) or Level 2 masking noise (Tritone, $SD = 0.53$; Perfect 5th, $SD = 0.82$). The results of the ANOVA and subsequent pairwise comparisons are summarized in Tables 1 and 2 respectively. Interval had a significant effect on pleasantness rating ($F_{(1, 12)} = 15.96$, $p < 0.001$, $\eta^2 = 0.12$) with the Perfect 5th ($M = -0.75$) on average being rated as more pleasant than the Tritone ($M = -1.41$; $t_{(38)} = 5.85$, $p < 0.001$). Masking noise also had a significant effect ($\epsilon = 0.57$, $F_{(1.14, 13.68)} = 40.48$, $p < 0.001$, $\eta^2 = 0.51$). In Bonferroni corrected t -tests ($\alpha = 0.017$) pleasantness ratings in the No Noise condition ($M = 0.13$) were significantly greater than Level 1 ($M = -1.30$; $t_{(25)} = 8.18$, $p < 0.001$) and Level 2 conditions ($M = -2.07$; $t_{(25)} = 8.84$, $p < 0.001$). Pleasantness ratings in the Level 1 ($M = -1.30$) condition were significantly greater than in the Level 2 condition ($M = -2.07$; $t_{(25)} = 7.05$, $p < 0.001$).

The effect of masking noise significantly interacted with the effect of interval ($\epsilon = 0.79$, $F_{(1.58, 18.96)} = 4.34$, $p = 0.04$, $\eta^2 = 0.01$; Fig. 2B). To explore the interaction further, consonance preference scores were calculated (Fig. 2C). In Bonferroni corrected paired t -tests ($\alpha = 0.017$) consonance preference in the No Noise condition ($M = 0.92$) was significantly greater than in the Level 2 condition ($M = 0.33$; $t_{(12)} = 2.48$, $p = 0.01$) but not the Level 1 condition ($M = 0.63$; $t_{(12)} = 1.40$, $p = 0.09$). Consonance preference scores in the Level 1 and Level 2 condition were not significantly different ($t_{(12)} = 2.07$, $p = 0.03$).

3.2. The effect of high-pass masking noise on the FFR to low-frequency dyads

Fig. 3 displays the average power spectra of the FFR_{RAW} to the Tritone (Fig. 3A–C) and the Perfect 5th (Fig. 3D–F). The power spectrum of the FFR_{RAW} to the consonant Perfect 5th in the No Noise condition (Fig. 3D) contains clearly defined peaks, at frequencies both present and not present in the stimulus. As noted elsewhere, the distortion products in the spectrum serve to enhance the overall harmonicity (Bones et al., 2014; Lee et al., 2009). The FFR_{RAW} to the dissonant Tritone in the same condition

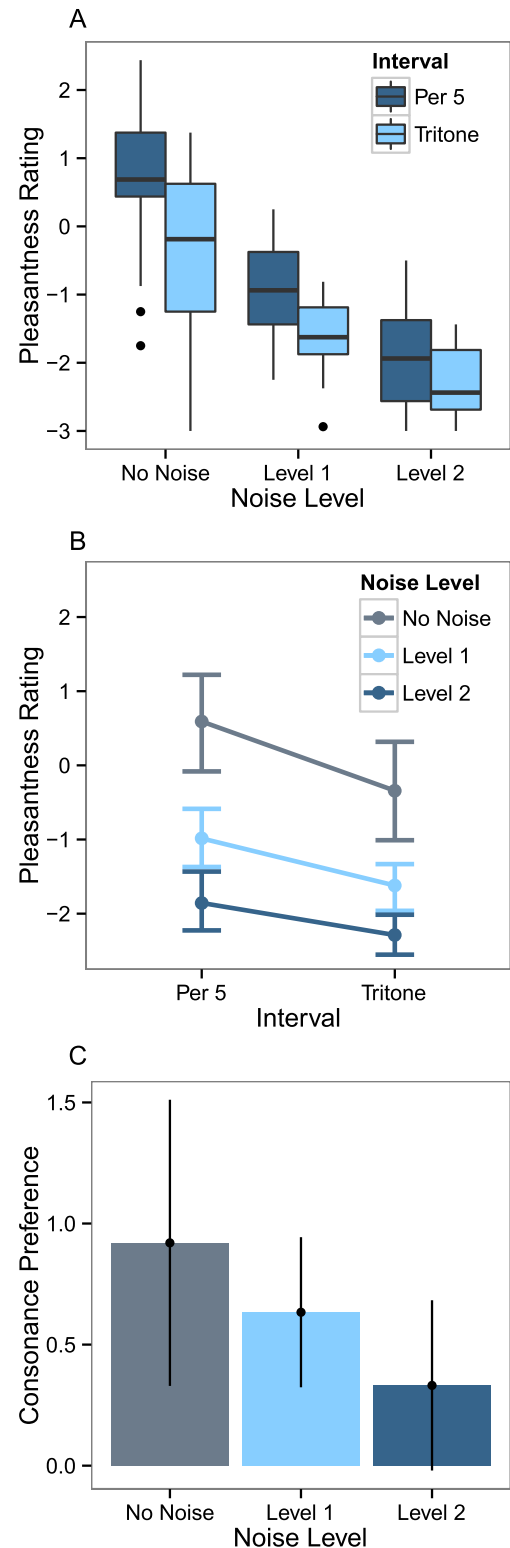


Fig. 2. A. Pleasantness ratings as a function of noise level, grouped by interval. Filled circles indicate outliers (greater than $\pm 1.5 \times$ inter-quartile range). B. Pleasantness ratings as a function of interval, grouped by noise level. Error bars represent 95% confidence intervals. C. Consonance preference calculated from pleasantness ratings as a function of noise level. Error bars represent 95% confidence intervals.

(Fig. 3A) also contains peaks in the spectrum corresponding to the frequency components present in the stimulus although they are not as clearly defined and are lower in amplitude than is the case in the FFR_{RAW} to the Perfect 5th.

Table 1
Summary of ANOVA effects and interactions.

		Noise	Interval	Noise \times interval
Pleasantness ratings	<i>F</i>	40.48	15.96	4.34
	<i>p</i>	<0.001**	<0.001**	0.04*
Neural harmonic salience	<i>F</i>	9.52	14.90	13.26
	<i>p</i>	0.01*	0.03*	0.004**

* indicates significance at $p < 0.05$, ** indicates significance at $p < 0.01$.

With masking noise added to the dyads (Level 1, middle row; and Level 2, bottom row) the representation of the stimuli in the spectra of the FFR_{RAW} to the Perfect 5th and the Tritone are less well defined. In the case of the Perfect 5th the large peaks corresponding to frequencies in the stimulus are noticeably reduced in amplitude, and many of the distortion products are no longer above the background EEG noise floor.

Neural harmonic salience scores are plotted in Fig. 4A. As for pleasantness ratings, for both intervals variance in neural harmonic salience was greater in the No Noise condition (Tritone, $SD = 0.32$; Perfect 5th, $SD = 1.24$) than with Level 1 (Tritone, $SD = 0.18$; Perfect 5th, $SD = 0.37$) and Level 2 masking noise (Tritone, $SD = 0.19$; Perfect 5th, $SD = 0.27$). The results of the ANOVA and subsequent post-hoc tests are also summarized in Tables 1 and 2 respectively. Musical interval had a significant effect on neural harmonic salience ($F_{(1, 9)} = 14.90$, $p = 0.03$, $\eta^2 = 0.191$) with the Perfect 5th having greater neural harmonic salience on average ($M = 1.13$) than the Tritone ($M = 0.61$, $t_{(29)} = 3.66$, $p < 0.001$). Adding high-pass masking noise also had a significant effect ($\epsilon = 0.53$, $F_{(1.06, 9.54)} = 9.52$, $p = 0.01$, $\eta^2 = 0.234$). In Bonferroni corrected paired *t*-tests ($\alpha = 0.017$) neural harmonic salience in the No Noise condition ($M = 1.29$) was greater than in the Level 1 ($M = 0.67$; $t_{(19)} = 3.05$, $p = 0.007$) and Level 2 condition ($M = 0.65$; $t_{(19)} = 2.96$, $p = 0.008$). Harmonic salience was not significantly different between the Level 1 and Level 2 condition ($t_{(19)} = 0.28$, $p = 0.79$). Masking noise and interval interacted ($\epsilon = 0.54$, $F_{(1.08, 9.72)} = 13.26$, $p = 0.004$, $\eta^2 = 0.180$; Fig. 4B).

NCI is a measure of how much more salient the harmonicity of consonant dyads are than dissonant dyads. The large difference in salience between the Perfect 5th and the Tritone in the No Noise condition but not the masking noise conditions seen in Fig. 4B is reflected in the NCI measures (Fig. 4C). In Bonferroni corrected paired *t*-tests ($\alpha = 0.017$) NCI scores in the No Noise condition ($M = 1.24$) were significantly greater than in the Level 1 ($M = 0.12$; $t_{(9)} = 3.96$, $p = 0.003$) and Level 2 conditions ($M = 0.21$, $t_{(9)} = 2.96$, $p = 0.008$). NCI in the Level 1 condition was not significantly different to the Level 2 condition ($t_{(9)} = -1.24$, $p = 0.88$).

3.3. The relation between NCI and individual preference for consonance for non-musicians

Individual consonance preference of young normal hearing non-musicians is plotted as a function of NCI for each masking noise condition in Fig. 5. NCI predicted consonance preference in the No Noise condition, with individuals with larger NCI scores also having

Table 2
Summary of pairwise comparisons.

	Pleasantness ratings		Neural harmonic salience	
	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>
Perfect 5th v Tritone	5.85	<0.001**	3.66	<0.001**
No Noise v Level 1	8.18	<0.001**	3.05	0.007**
No Noise v Level 2	8.84	<0.001**	2.96	0.008**
Level 1 v Level 2	7.05	<0.001**	0.28	0.79

* indicates significance at $p < 0.05$, ** indicates significance at $p < 0.01$.

a greater preference for consonance over dissonance ($r_{s(8)} = 0.61$, $p = 0.03$). The reduction in variance of both neural harmonic salience and pleasantness ratings when masking noise was added to the stimulus can be seen in the clustering of data points in the Level 1 and Level 2 conditions in the bottom left of the plot. Correlations between consonance preference and NCI in the Level 1 ($r_{s(8)} = 0.20$, $p = 0.29$) and Level 2 ($r_{s(8)} = -0.12$, $p = 0.63$) conditions were not significant.

Bones et al. (2014) found that NCI only predicted consonance preference when phase locking to the cochlear envelope was suppressed. In the current study however NCI predicted consonance preference without this component suppressed. To explore this further the spectra of the FFRs to the Perfect 5th in each polarity were added and subtracted. Fig. 6A–C displays the spectrum of the FFR_{RAW} (A), FFR_{ADD} (the response to the envelope enhanced, the response to the TFS suppressed; B), and FFR_{SUB} (the response to the envelope suppressed, the response to the TFS enhanced; C) to the Perfect 5th. Each FFR type to the Perfect 5th from Bones et al. is displayed in Fig. 6D–F for comparison. Bones et al. used dyads with a 130.81 Hz root note F0. They found that when consonant dyads were presented diotically they produced large distortion products corresponding to the difference tone of the F0 of the two notes ($F_2 - F_1$). For the Perfect 5th the interval note F0 was 196 Hz. The large difference tone of approximately 65 Hz corresponded to the F0 of the harmonic series of the FFR, serving to enhance the overall harmonicity of the spectrum (Fig. 6D–E). Suppressing the response to the envelope also suppressed the difference tone (Fig. 6F), suggesting that phase locking at this frequency was to the cochlear envelope resulting from interaction of the two notes (Bones et al., 2014; Gockel et al., 2012). In the current study, despite the FFR to the No Noise Perfect 5th stimulus containing distortion products at most harmonic frequencies up to 2500 Hz, the difference tone corresponding to the F0 (110 Hz) is lower in magnitude (Fig. 6A). When the envelope component is enhanced (Fig. 6B) the difference tone is of a greater magnitude. When masking noise was added to the stimulus (Fig. 3B–C) the difference tone had a greater magnitude than in the No Noise condition, and was greater in magnitude than the other components of the spectrum.

Correlations between NCI for each FFR type and consonance preference are plotted in Fig. 7. Greater consonance preference was associated with larger NCI scores for all FFR types (Raw, $r_{s(8)} = 0.61$, $p = 0.03$; Add, $r_{s(8)} = 0.50$, $p = 0.05$; Sub, $r_{s(8)} = 0.61$, $p = 0.03$).

3.4. Comparison of the effects of high-pass masking noise on the FFR and a model of the inner hair cell receptor-potential

To explore whether masking noise reduced harmonic salience of the FFR to the Perfect 5th by reducing power in the spectrum at harmonic peaks or by increasing power in the spectrum at non-harmonic frequencies (neural noise), a two-way repeated measures ANOVA of spectral power was performed. One factor of the ANOVA was frequency region with levels “harmonic peaks” (i.e. the power inside the bins used to calculate harmonic salience) and “neural noise” (i.e. the power of the spectrum between 0 and 2500 Hz outside of the bins used to calculate harmonic salience). Masking noise level was the other factor. Masking noise level was marginally non-significant after correcting for non-sphericity ($\epsilon = 0.52$, $F_{(1.04, 9.36)} = 4.11$, $p = 0.07$, $\eta^2 = 0.010$). Frequency region was a significant effect ($F_{(1, 19)} = 10.30$, $p = 0.01$, $\eta^2 = 0.213$). Moreover, frequency region interacted with masking noise level ($\epsilon = 0.51$, $F_{(1.02, 9.18)} = 5.41$, $p = 0.04$, $\eta^2 = 0.011$). The interaction between frequency and masking noise level can be seen in Fig. 8. In Bonferroni correct *t*-tests ($\alpha = 0.017$) the difference in power of harmonic peaks in the No Noise condition ($M = 2.51 \times 10^{-15} \text{ V}^2$) and the Level 1 ($M = 0.89 \times 10^{-15} \text{ V}^2$; $t_{(9)} = 2.19$, $p = 0.06$) and Level 2

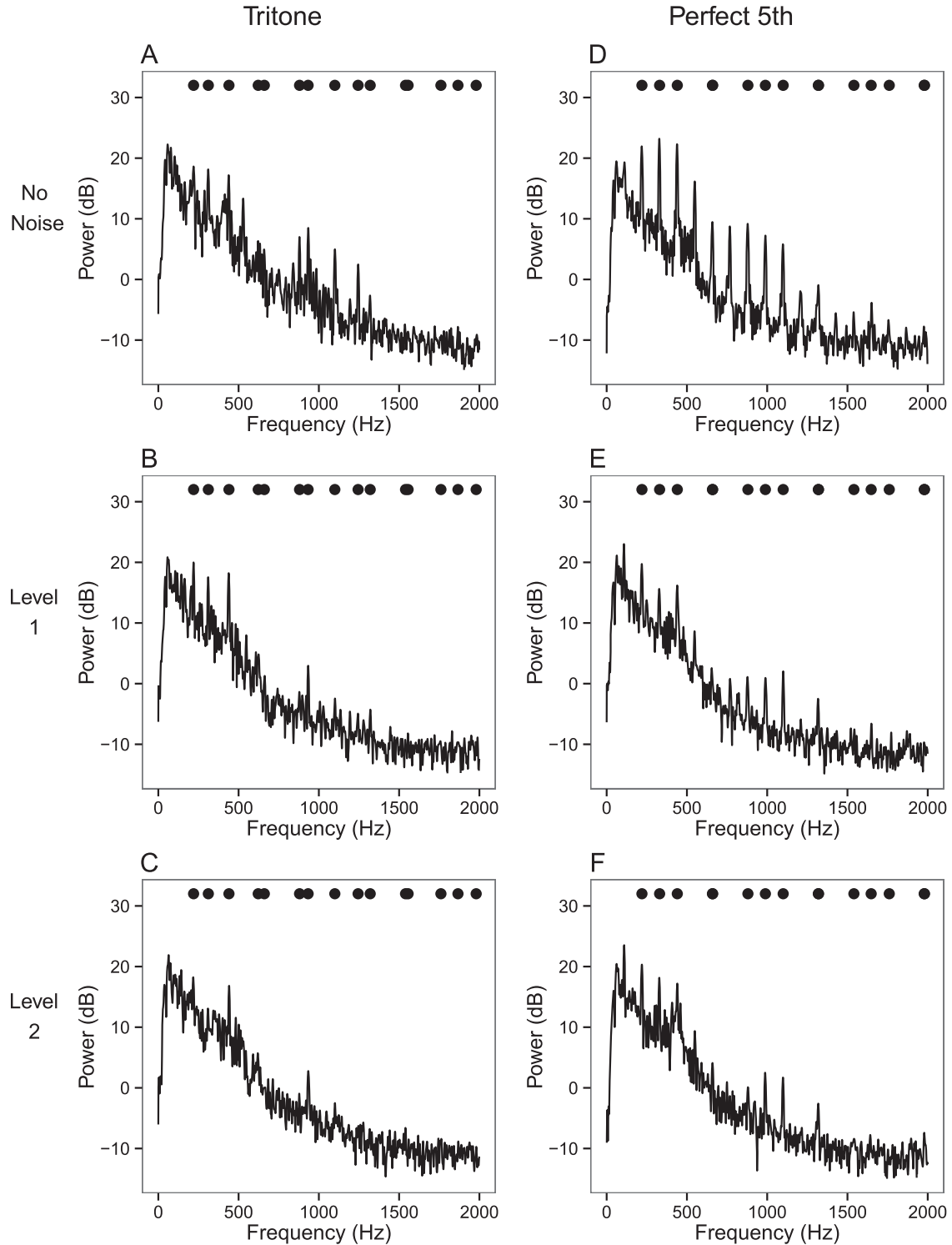


Fig. 3. The average spectra of the FFR_{RAW} recordings from the two polarities (the FFR as it was recorded) to the Tritone (A–C, left column) and the Perfect 5th (D–F, right column), with No Noise (top row), Level 1 noise (middle row), and Level 2 noise (bottom row). Black circles indicate the frequency components present in the stimulus. Power is expressed as dB referenced to 10^{-16} V^2 .

condition ($M = 0.88 \times 10^{-15} \text{ V}^2$; $t_{(9)} = 2.18$, $p = 0.06$) was not significant. Differences between neural noise in the No Noise ($M = 2.39 \times 10^{-16} \text{ V}^2$) condition and the Level 1 ($M = 2.93 \times 10^{-16} \text{ V}^2$; $t_{(9)} = -1.29$, $p = 0.23$) and Level 2 ($M = 2.93 \times 10^{-16} \text{ V}^2$; $t_{(9)} = 0.95$, $p = 0.37$) condition were also not significant.

The output of the IHC RP model to the Perfect 5th stimulus is displayed in Fig. 9. First consider the No Noise condition (Fig. 9A–D). The response of the LF model (containing frequency channels corresponding to the frequency content of the stimulus) contains a DC component (Fig. 9A), with frequency components corresponding to the stimulus and also the distortion products seen

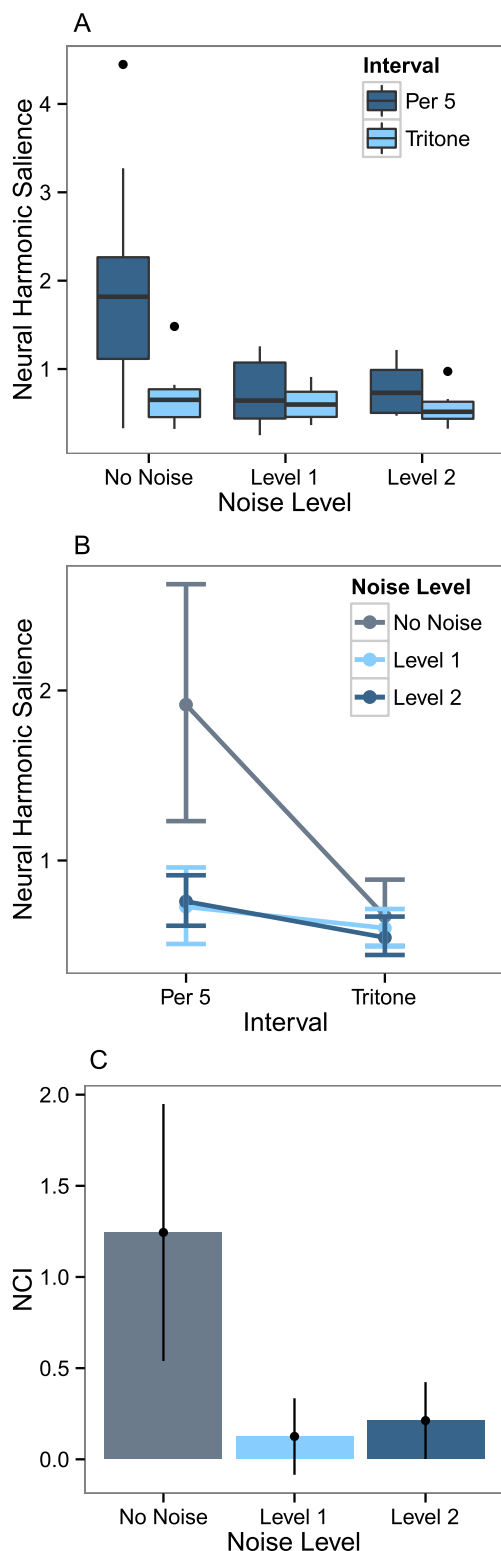


Fig. 4. A: Neural harmonic salience, a measure of how well the harmonicity of the stimulus is represented in the FFR_{RAW}, as a function of noise level, grouped by interval. Filled circles indicate outliers (values greater than $\pm 1.5 \times$ inter-quartile range). Error bars represent 95% confidence intervals. B: Neural harmonic salience as a function of interval, grouped by noise level. Error bars represent 95% confidence intervals. C: NCI calculated from neural harmonic salience values, as a function of noise level. Error bars represent 95% confidence intervals.

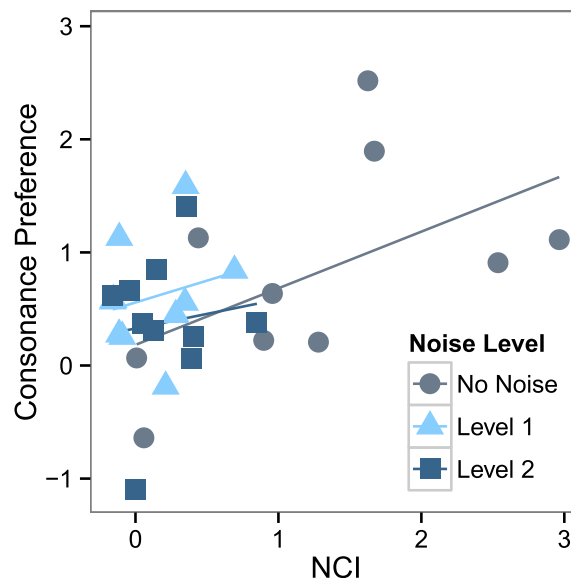


Fig. 5. Consonance preference, calculated from pleasantness ratings, of young normal hearing non-musicians as a function of NCI, calculated from neural harmonic salience, for each noise level.

in the FFR data (Fig. 9B). The output of the HF model, representing the response of IHCs with CFs higher than the frequency content of the stimulus is lower in amplitude, with a smaller DC component (Fig. 9C). The IHC RP of the HF model also contains distortion products which serve to reinforce the harmonicity of the response (Fig. 9D), although they are lower in amplitude than those in the output of the LF model.

When Level 2 masking noise is added to the stimulus the response of the LF model is virtually identical to the No Noise condition (Fig. 9E–F). However, the response of the HF model to Level 2 masking noise (Fig. 9G) is markedly different to the response of the HF model to No Noise (Fig. 9C). The waveform has a DC component that is larger than the AC component, the harmonic frequency components are reduced in amplitude and the background noise floor is increased in amplitude (Fig. 9H).

The power of the harmonic peaks and neural noise floor in the output of each model in each condition is summarized in Fig. 10. In the LF model, the mean power of harmonic peaks remains virtually constant between the No Noise ($M = 3.49 \times 10^{-8} \text{ V}^2$) and the Level 2 ($M = 3.46 \times 10^{-8} \text{ V}^2$) conditions, whereas, similar to the FFR data, the mean power of the harmonic peaks in the output of the HF model reduces from No Noise ($M = 0.99 \times 10^{-8} \text{ V}^2$) to Level 2 masking noise ($M = 0.47 \times 10^{-8} \text{ V}^2$). The power of the neural noise floor also remains virtually constant between No Noise ($M = 5.58 \times 10^{-10} \text{ V}^2$) and Level 2 masking noise ($M = 5.68 \times 10^{-10} \text{ V}^2$) in the LF model. However, the background noise floor of the HF model increases from No Noise ($M = 1.70 \times 10^{-10} \text{ V}^2$) to Level 2 masking noise ($M = 4.75 \times 10^{-10} \text{ V}^2$).

4. Discussion

4.1. Neural harmonic salience of the FFR and musical consonance perception of non-musicians

The results of this study provide further evidence that the salience of subcortical temporal coding of the harmonicity of consonant dyads relative to dissonant dyads as represented by the FFR predicts individual preference for consonance over dissonance. One of the aims of the current study was to determine whether the relation between NCI (a measure of the salience of the harmonicity

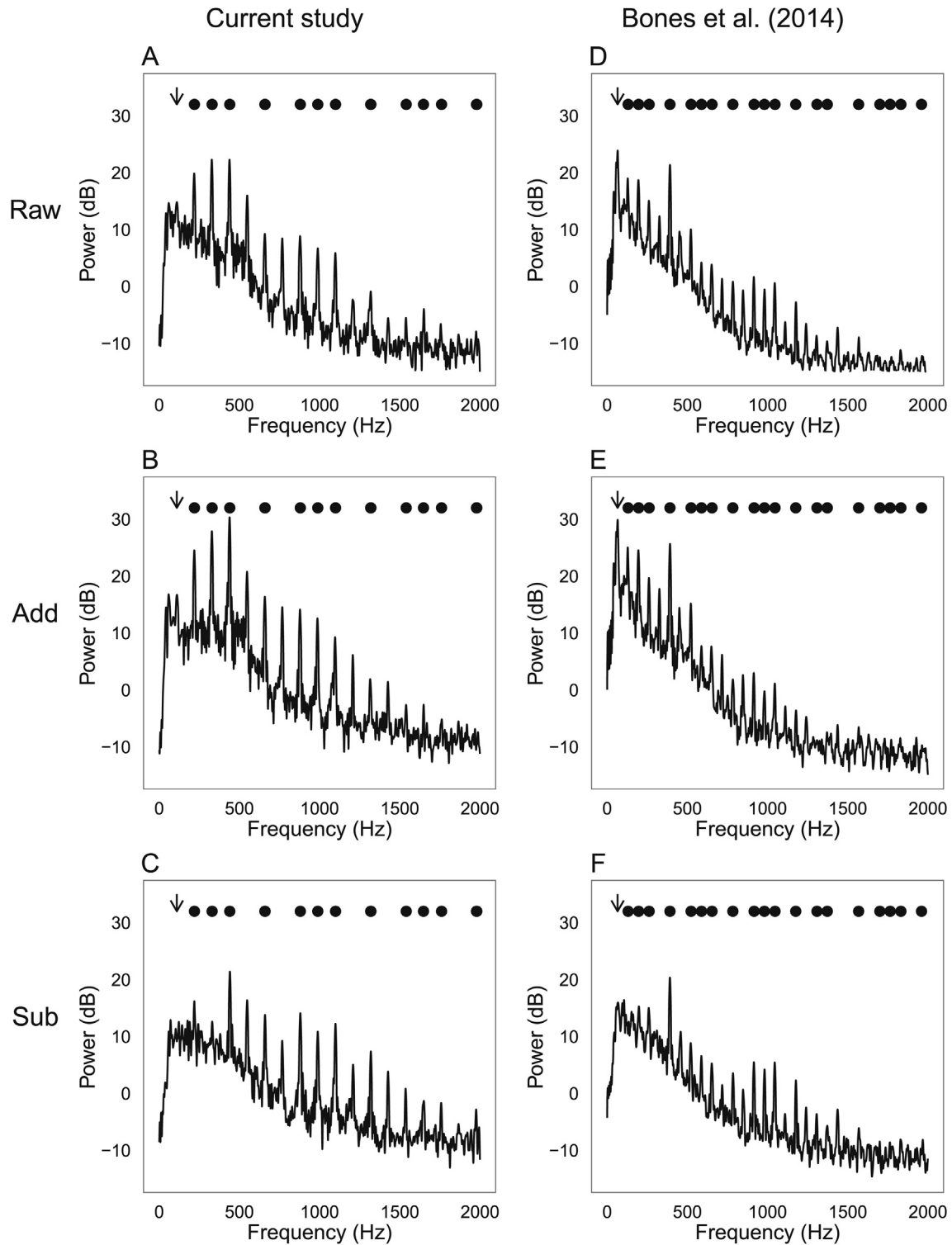


Fig. 6. Average spectra of the FFR_{RAW} (the FFR as it was recorded, with the spectra of the two polarities averaged; top), FFR_{ADD} (the response to the two stimulus polarities added; middle), and FFR_{SUB} (the response to the two polarities subtracted; bottom) to the Perfect 5th from the current study (A–C) and Bones et al. (2014; D–F). Power is expressed as dB referenced to 10^{-16} V^2 . Black circles indicate frequencies present in the stimulus, arrows indicate the implied F0 of the harmonic series of the spectrum of the FFR.

of the neural response to consonant relative to dissonant dyads) and consonance preference reported elsewhere for young normal hearing listeners with a range of musical experience (Bones et al., 2014) also occurs for young normal hearing listeners with no musical experience. The current study found a significant correlation between NCI and consonance preference for young normal

hearing listeners with no musical experience. To the authors' knowledge this is the first time that this has been demonstrated.

As has been shown previously (Bones et al., 2014; Lee et al., 2009), the FFR to a consonant dyad in the current study contained multiple distortion products which enhanced the overall harmonicity. Lee et al. (2009) found that distortion products had

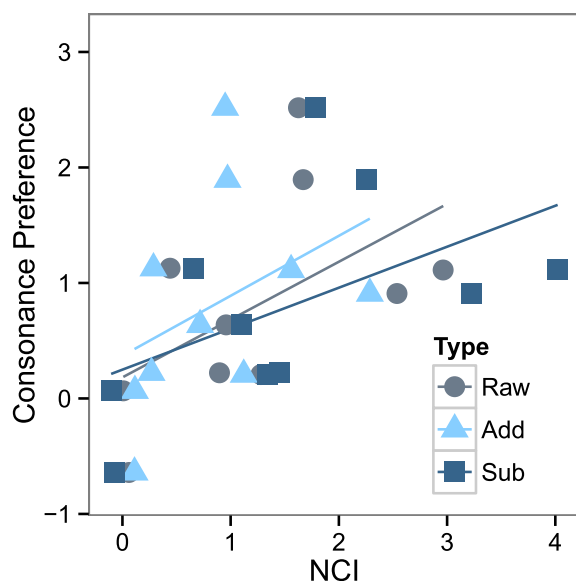


Fig. 7. Consonance preference, calculated from pleasantness ratings, as a function of NCI, calculated from the neural harmonic salience of the FFR_{RAW} (the FFR as it was recorded, with the spectra of the two polarities averaged), FFR_{ADD} (the response to the two stimulus polarities added), and FFR_{SUB} (the response to the two polarities subtracted).

greater magnitude in the FFR of musicians compared to non-musicians. Lerud et al. (2014) suggest that this might be due to greater synaptic efficiency in musicians, leading to more nonlinear processing. The data presented here demonstrate that the nonlinearity of the auditory system of non-musicians also produces large distortion products which serve to enhance the harmonicity of the FFR to consonant dyads. Surprisingly however, the large difference tones in the FFR to consonant dyads reported by Bones et al. (2014) were not found in the No Noise condition of the current study. One possibility is that this is a consequence of the different F0s: the F0s of the root and interval note of the Perfect 5th dyad used in the current study and by Bones et al. were 220 and 440, and 130.81 and 196 Hz respectively. The smaller sample-size of the current study compared to the sample size of Bones et al. ($N = 10, 19$) may also be a contributing factor.

Bones et al. found that the FFR to consonant dyads presented diotically had greater harmonic salience than when presented dichotically, due to the difference tone produced by monaural interactions in the diotic condition. This resulted in larger NCI scores in the diotic condition, and coincided with higher pleasantness ratings for consonant dyads and therefore greater consonance preference. NCI only predicted consonance preference when the envelope and difference tone were suppressed (FFR_{SUB}). In the current study NCI calculated from FFR_{RAW} was predictive of consonance preference. Whether suppression of the envelope component was not necessary due to the absence of the large difference tone previously reported will need further investigation.

The strong interaction between the effects of masking noise and interval on harmonic salience were not found for pleasantness ratings. Ratings for both the Tritone and the Perfect 5th declined incrementally with Level 1 and Level 2 masking noise, with both dyads being rated as being more pleasant in the Level 1 condition than in the Level 2 condition. It is likely that the reduction in pleasantness ratings when masking noise was added to the dyads is representative of the stimuli becoming increasingly irritating in a way unrelated to the perception of consonance *per se*.

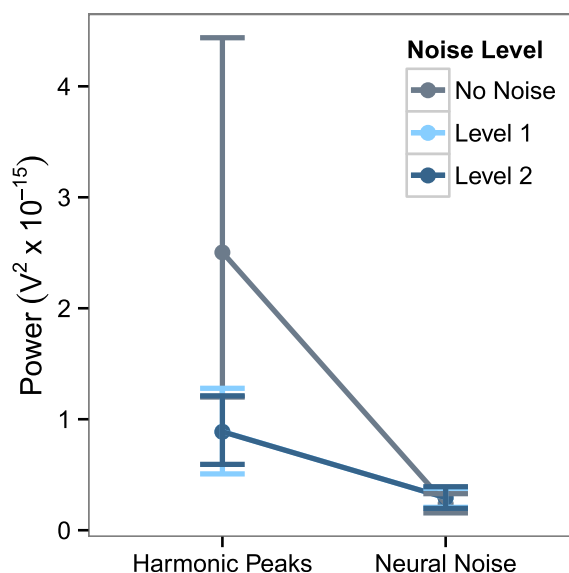


Fig. 8. The average power of the FFR_{RAW} to the Perfect 5th at frequencies of the spectrum of that contribute to its harmonicity (Harmonic Peaks) and that do not (Neural Noise) for each noise level. Error bars represent 95% confidence intervals.

4.2. The FFR to low-frequency musical dyads is likely to be partly generated in the basal region of the cochlea

The results of the current study suggest that the FFR to the low-frequency Perfect 5th dyad was partly generated by a region of the cochlea tuned to frequencies above the dyad. The addition of high-pass masking noise reduced the neural harmonic salience of the FFR. This effect was driven by a reduction in the amplitude of harmonic components of the FFR, rather than an overall increase in the background EEG noise floor, implying a nonlinear process.

Previous work has demonstrated that high-frequency masking noise may reduce the amplitude of the cochlear microphonic to a low-frequency tone (Zhang, 2014). In the current study we demonstrate that the reduction in amplitude of the FFR to a low-frequency musical dyad by high-frequency masking noise can be accounted for by a model of IHC RP. With no masking noise added to the stimulus the spectra of both the LF and HF model output (Fig. 9B, F) are similar to the FFR spectrum in the same condition (Fig. 3D); both spectra contain frequencies of the stimulus and distortion products, indicating that processing in both the low- and the high-frequency pathway up to the output of the IHC is sufficiently nonlinear to generate additional harmonic frequencies. This is consistent with a saturating IHC response (Dallos, 1986; Patuzzi and Sellick, 1983), and demonstrates that even though the BM response to the low-frequency dyad at a place tuned to a higher-frequency is likely to be linear, the IHC response is sufficiently nonlinear to produce distortion products. It should be noted that distortion products in the FFR may also represent further nonlinear processing beyond the IHC (e.g. see Lins et al., 1995, p. 3059, Fig. 9).

Adding masking noise to the stimulus in the LF model has no effect because the frequency range of the masking noise is above the frequency range of the channels included in the model; in the HF model the masking noise has a clear effect on the IHC RP waveform (Fig. 6G) and corresponding spectrum (Fig. 6H). The harmonic peaks of the spectrum are reduced from a mean amplitude of 0.99 to $0.47 \times 10^{-8} \text{ V}^2$. The ratio of this reduction in amplitude (0.47) was not significantly different to the mean ratio of the reduction of the harmonic peaks in the FFR spectra (0.55 ; $r = -0.28$, $V = 18$, $p = 0.38$). The reduction in the amplitude of harmonic frequency components in the FFR to low-frequency dyads

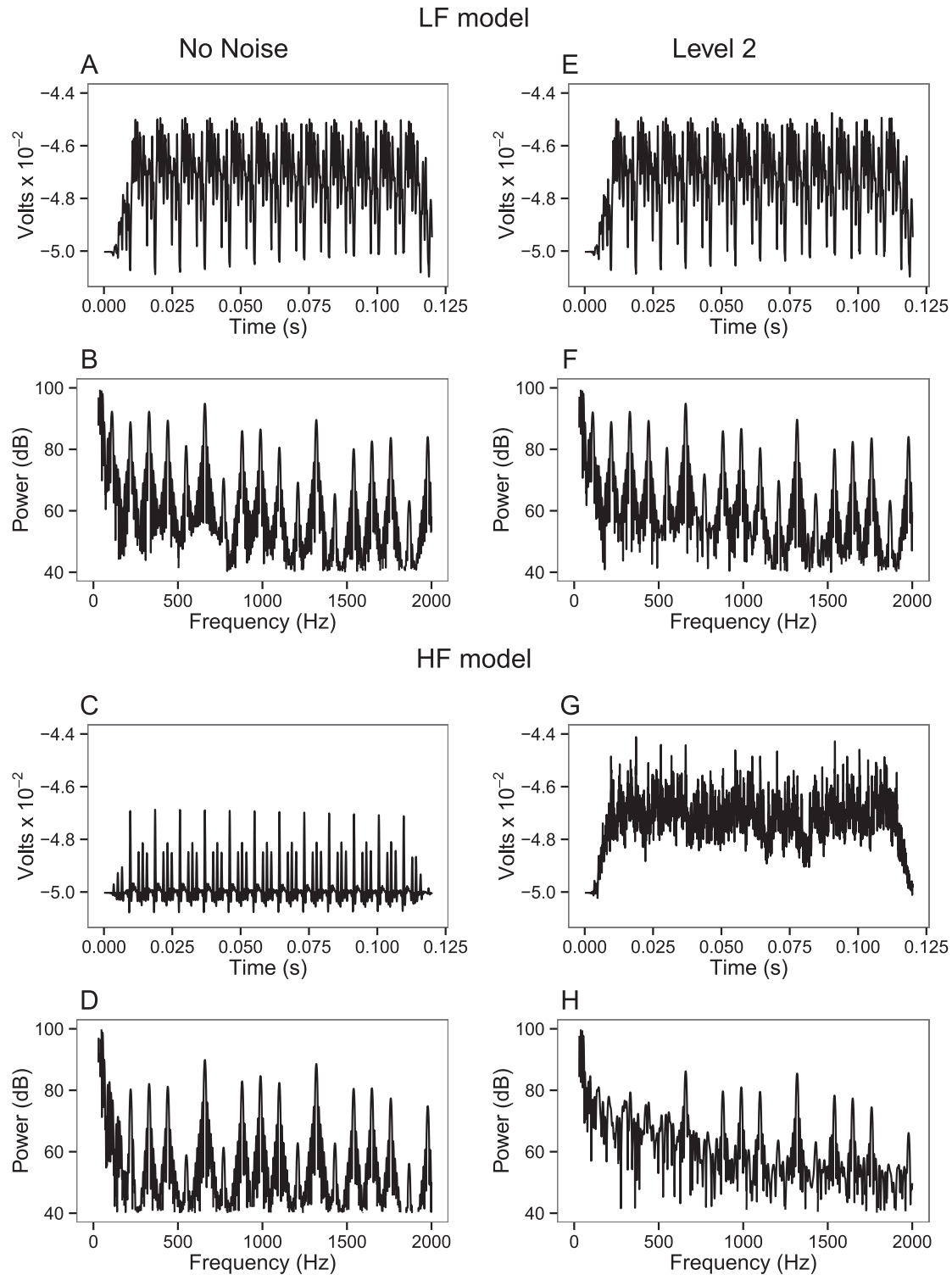


Fig. 9. The output of the IHC RP of the LF model consisting of only low-frequency channels (waveforms top row, spectra second row) and HF model consisting of only high-frequency channels (waveforms third row, spectra fourth row) in response to the Perfect 5th stimulus with No Noise (A–D, left column) and Level 2 noise (E–H, right column). Power is expressed as dB referenced to 10^{-16} V^2 .

when high-pass masking noise is added can therefore be accounted for by the saturating response of IHCs in the basal portion of the cochlea.

The ratio of the increase in noise floor in the output of the HF model from No Noise to Level 2 masking noise (2.79) is greater

than the mean ratio of the increase in background EEG noise of the FFR in the same condition (1.66). Although this difference is not statistically significant ($r = -0.54$, $V = 45$, $p = 0.08$), it should be noted that the FFR represents the integration of the evoked response with spontaneous neural activity and the electrical

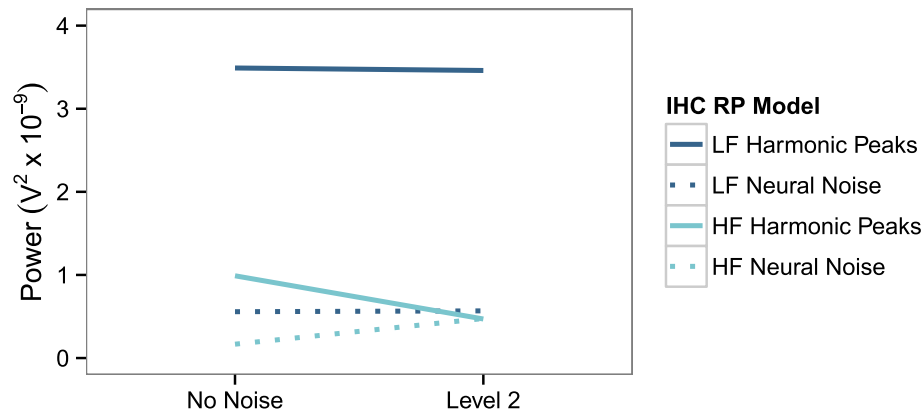


Fig. 10. The average power at frequencies that contribute to the harmonicity of the spectrum (Harmonic Peaks) and at frequencies that do not (Neural Noise) in the output of the IHC RP models for No Noise and Level 2 noise.

potentials created by muscular activity, and therefore has a background noise floor even in the No Noise condition. It is possible therefore that the FFR is not sensitive to the increase in the noise floor of the IHC RP in response to Level 2 masking noise.

5. Conclusions

- 1) The salience of the harmonicity of consonant dyads relative to dissonant dyads in the FFR is predictive of individual consonance preference in young normal hearing non-musicians.
- 2) The harmonic salience of the FFR to low-frequency dyads is reduced by high-pass masking noise above the frequency range of the dyads.
- 3) The reduction in harmonic salience due to the addition of the noise is the result of a reduction in amplitude of the harmonic components of the FFR. This can be accounted for by saturating high-frequency IHC RPs, suggesting that the FFR to low-frequency dyads is at least partly generated in the basal region of the cochlea.

References

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011a. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011b. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn.* 77, 1–10.

Bidelman, G.M., Krishnan, A., 2009. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171.

Bledsoe Jr., S.C., Moushegian, G., 1980. The 500 Hz frequency-following potential in kangaroo rat: an evaluation with masking noise. *Electroencephalogr. Clin. Neurophysiol.* 48, 654–663.

Bones, O., Hopkins, K., Krishnan, A., Plack, C.J., 2014. Phase locked neural activity in the human brainstem predicts preference for musical consonance. *Neuropsychologia* 58C, 23–32.

Brugge, J.F., Anderson, D.J., Hind, J.E., Rose, J.E., 1969. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J. Neurophysiol.* 32, 386–401.

Carcagno, S., Plack, C.J., 2011. Subcortical plasticity following perceptual learning in a pitch discrimination task. *J. Assoc. Res. Otolaryngol.* 12, 89–100.

Clinard, C.G., Tremblay, K.L., Krishnan, A.R., 2010. Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hear. Res.* 264, 48–55.

Dallos, P., 1986. Neurobiology of cochlear inner and outer hair cells: intracellular recordings. *Hear. Res.* 22, 185–198.

Dau, T., 2003. The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J. Acoust. Soc. Am.* 113, 936–950.

Dau, T., Wegner, O., Mellert, V., Kollmeier, B., 2000. Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. *J. Acoust. Soc. Am.* 107, 1530–1540.

Davis, R., Britt, R., 1984. Analysis of the frequency following response in the cat. *Hear. Res.* 15, 29–37.

Davis, H., Hirsh, S.K., 1976. The audiometric utility of brain stem responses to low-frequency sounds. *Audiology* 15, 181–195.

de Boer, E., 1977. Cochlear events and whole-nerve action potentials. *Audiology* 16, 380–388.

Gardi, J., Merzenich, M., 1979. Effect of high-pass noise on the scalp-recorded frequency following response (FFR) in humans and cats. *J. Acoust. Soc. Am.* 65, 1491–1500.

Geisler, C.D., Rhode, W.S., Kennedy, D.T., 1974. Responses to tonal stimuli of single auditory-nerve fibers and their relationship to basilar-membrane motion in squirrel-monkey. *J. Neurophysiol.* 37, 1156–1172.

Goblick Jr., T.J., Pfeiffer, R.R., 1969. Time-domain measurements of cochlear non-linearities using combination click stimuli. *J. Acoust. Soc. Am.* 46, 924–938.

Gockel, H.E., Carlyon, R.P., Mehta, A., Plack, C.J., 2011. The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *J. Assoc. Res. Otolaryngol.* 12, 767–782.

Gockel, H.E., Farooq, R., Muhammed, L., Plack, C.J., Carlyon, R.P., 2012. Differences between psychoacoustic and frequency following response measures of distortion tone level and masking. *J. Acoust. Soc. Am.* 132, 2524–2535.

Heinz, M.G., Zhang, X., Bruce, I.C., Carney, L.H., 2001. Auditory nerve model for predicting performance limits of normal and impaired listeners. *Acoust. Res. Lett. Online* 2, 91–96.

Huis in't Veld, F., Osterhammel, P., Terkildsen, K., 1977. The frequency selectivity of the 500 Hz frequency following response. *Scand. Audiol.* 6, 35–42.

Kiang, N.Y., Moxon, E.C., 1974. Tails of tuning curves of auditory-nerve fibers. *J. Acoust. Soc. Am.* 55, 620–630.

Krishnan, A., 2007. Frequency-following response. In: Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins, pp. 313–333.

Krishnan, A., Bidelman, G.M., Smalt, C.J., Ananthakrishnan, S., Gandour, J.T., 2012. Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia* 50, 2849–2859.

Krishnan, A., Plack, C.J., 2011. Neural encoding in the human brainstem relevant to the pitch of complex tones. *Hear. Res.* 275, 110–119.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.* 25, 161–168.

Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.

Lerud, K.D., Almonte, F.V., Kim, J.C., Large, E.W., 2014. Mode-locking neurodynamics predict human auditory brainstem responses to musical intervals. *Hear. Res.* 308, 41–49.

Lins, O.G., Picton, P.E., Picton, T.W., Champagne, S.C., Durieux-Smith, A., 1995. Auditory steady-state responses to tones amplitude-modulated at 80–110 Hz. *J. Acoust. Soc. Am.* 97, 3051–3063.

Marmel, F., Linley, D., Carlyon, R.P., Gockel, H.E., Hopkins, K., Plack, C.J., 2013. Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *J. Assoc. Res. Otolaryngol.* 14.

Marsh, J.T., Smith, J.C., Worden, F.G., 1972. Receptor and neural responses in auditory masking of low frequency tones. *Electroencephalogr. Clin. Neurophysiol.* 32, 63–74.

McDermott, J.H., Lehr, A.J., Oxenham, A.J., 2010. Individual differences reveal the basis of consonance. *Curr. Biol.* 20, 1035–1041.

Meddis, R., O'Mard, L.P., Lopez-Poveda, E.A., 2001. A computational algorithm for computing nonlinear auditory frequency selectivity. *J. Acoust. Soc. Am.* 109, 2852–2861.

- Moushegian, G., Rupert, A.L., Stillman, R.D., 1978. Evaluation of frequency-following potentials in man: masking and clinical studies. *Electroencephalogr. Clin. Neurophysiol.* 45, 711–718.
- Patuzzi, R., Sellick, P.M., 1983. A comparison between basilar membrane and inner hair cell receptor potential input-output functions in the guinea pig cochlea. *J. Acoust. Soc. Am.* 74, 1734–1741.
- Pressnitzer, D., Patterson, R.D., Krumbholz, K., 2001. The lower limit of melodic pitch. *J. Acoust. Soc. Am.* 109, 2074–2084.
- Rose, J.E., Hind, J.E., Anderson, D.J., Brugge, J.F., 1971. Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey. *J. Neurophysiol.* 34, 685–699.
- Shamma, S.A., Chadwick, R.S., Wilbur, W.J., Morrish, K.A., Rinzel, J., 1986. A biophysical model of cochlear processing: intensity dependence of pure tone responses. *J. Acoust. Soc. Am.* 80, 133–145.
- Shore, S.E., Nuttall, A.L., 1985. High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts. *J. Acoust. Soc. Am.* 78, 1286–1295.
- Sumner, C.J., Lopez-Poveda, E.A., O'Mard, L.P., Meddis, R., 2002. A revised model of the inner-hair cell and auditory-nerve complex. *J. Acoust. Soc. Am.* 111, 2178–2188.
- Sumner, C.J., O'Mard, L.P., Lopez-Poveda, E.A., Meddis, R., 2003. A nonlinear filter-bank model of the guinea-pig cochlear nerve: rate responses. *J. Acoust. Soc. Am.* 113, 3264–3274.
- Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Yamada, O., Yamane, H., Kadera, K., 1977. Simultaneous recordings of the brain stem response and the frequency-following response to low-frequency tone. *Electroencephalogr. Clin. Neurophysiol.* 43, 362–370.
- Zhang, M., 2014. Effects of the intensity of masking noise on ear canal recorded low-frequency cochlear microphonic waveforms in normal hearing subjects. *Hear. Res.* 313, 9–17.